A study of the universal threshold in the ℓ_1 recovery by statistical mechanics

Koujin Takeda
Department of Computational Intelligence and
Systems Science
Tokyo Institute of Technology
Yokohama 226-8502, Japan
E-mail: takeda@sp.dis.titech.ac.jp

Yoshiyuki Kabashima
Department of Computational Intelligence and
Systems Science
Tokyo Institute of Technology
Yokohama 226-8502, Japan
E-mail: kaba@dis.titech.ac.jp

Abstract—We discuss the universality of the ℓ_1 recovery threshold in compressed sensing. Previous studies in the fields of statistical mechanics and random matrix integration have shown that ℓ_1 recovery under a random matrix with orthogonal symmetry has a universal threshold. This indicates that the threshold of ℓ_1 recovery under a non-orthogonal random matrix differs from the universal one. Taking this into account, we use a simple random matrix without orthogonal symmetry, where the random entries are not independent, and show analytically that the threshold of ℓ_1 recovery for such a matrix does not coincide with the universal one. The results of an extensive numerical experiment are in good agreement with the analytical results, which validates our methodology. Though our analysis is based on replica heuristics in statistical mechanics and is not rigorous, the findings nevertheless support the fact that the universality of the threshold is strongly related to the symmetry of the random matrix.

I. INTRODUCTION

Compressed sensing is nowadays one of the main topics in information science, where the sparsity of the signal plays an essential role. Compressed sensing has been intensively investigated from the theoretical point of view, and its application has been attempted in various fields of engineering.

We start with the basic ℓ_1-norm recovery problem proposed and analyzed elsewhere [1], [2], [3]. The sensing process in this problem is described by a linear equation:

$$y = Fx^0. \qquad (1)$$

(Bold letters denote vectors and matrices.) $x^0 \in \mathbb{R}^N$ is the input signal vector, $y \in \mathbb{R}^P$ is a P-dimensional observed signal vector, and F is a P-by-N sensing matrix. We assume that the entries in F are randomly generated. We define compression rate $\alpha := P/N < 1$, which is needed for the discussion of recovery performance. We also assume that input x^0 is also random and that it is drawn from a sparse distribution:

$$P(x_i^0) = (1-\rho)\delta(x_i^0) + \rho\widetilde{P}(x_i^0). \qquad (2)$$

Parameter ρ is the density of non-zero inputs, which is needed in the following. The distribution of nonzero entry $\widetilde{P}(x_i^0)$ can be set arbitrarily in principle. Here we set it as Gaussian with zero mean and unit variance.

Within this framework, we consider the ℓ_p-norm minimization problem with constraint

$$\text{minimize } \|x\|_p \quad \text{subject to} \quad y(= Fx^0) = Fx, \qquad (3)$$

where $\|x\|_p = \lim_{\epsilon \to +0} \sum_i |x_i|^{p+\epsilon}$. In particular we focus on the ℓ_1 problem ($p = 1$). This equation offers an algorithm of recovery for original input x^0, which we call ℓ_1-norm reconstruction. The problem discussed throughout this paper is a basic question: in which case does the solution vector of (3) coincide with original input x_0?

Under the ansatz of random sensing matrix F, the performance of ℓ_1-norm recovery has been evaluated using various approaches. One study using the restricted isometry property [1] in conjunction with the large deviation theory of random matrix spectral edge [3] showed that there is a perfect recovery region on the (α, ρ) plane. Another study using analysis of random polytope projection obtained a typical reconstruction threshold [4], [5], [6] that is in excellent agreement with the boundary between the success and failure regions obtained in an ℓ_1-norm reconstruction experiment. This typical threshold (termed weak threshold elsewhere [4], [5], [6]) can also be obtained by statistical mechanical analysis based on the replica method [7], which yields exactly the same analytical expression for the recovery threshold.

Though the rigorousness of the replica method has not yet been proven and this method is still a heuristics, it has a significant advantage: using the replica method, we can analyze problems more general than the basic ℓ_1-norm problem. For example, we previously used it to analyze the correlated sensing problem [8], [9]. For other generalizations, see the references in [9].

In this article we focus on the universality of the ℓ_1-norm recovery threshold discussed in [10]. From the statistical mechanical point of view, this universality can be comprehended from rotational symmetry in the matrix integration approach as analyzed and elucidated elsewhere [7], and with this knowledge we can construct a model that breaks such symmetry. In the following we propose a model of symmetry breaking by introducing a blockwise-correlated random sensing matrix and give an analytical expression for the recovery threshold. Using this expression, we can qualitatively trace the deformation

of the universal recovery threshold by introducing random matrix correlation. We also report the results of a numerical experiment for which the results are in excellent agreement with those of the proposed model.

This article is organized as follows. First we give an overview of statistical mechanical analysis using a replica method. Next we address the relationship between this analysis and threshold universality from the perspective of matrix integration. Then, as an example of how statistical mechanics can be used to analyze problems more general than the basic ℓ_1-norm problem, we use it to investigate the deformed problem of an i.i.d. random matrix. We next present a blockwise model for observing in detail the deviation from universality. Then we describe the numerical experiment we conducted to verify the results of the replica method, which lacks rigorousness. We conclude with a summary of the key points and a short discussion.

II. OVERVIEW OF STATISTICAL MECHANICAL ANALYSIS

We start with an outline of statistical mechanical analysis, as proposed elsewhere [7]. We focus on the basic model, where each entry in matrix $\boldsymbol{F}$ is drawn from a Gaussian distribution with zero mean and variance N^{-1}. For the moment we do not restrict ourselves to the ℓ_1-norm problem and consider instead the ℓ_p-norm problem.

The first step of the analysis is to define quantity C_p:

$$
\begin{aligned}
C_p \quad := \quad & -\lim_{\beta \to \infty} \lim_{N \to \infty} \frac{1}{\beta N} \\
& \times \ln \int d\boldsymbol{x} \exp(-\beta \|\boldsymbol{x}\|_p) \delta(\boldsymbol{F}(\boldsymbol{x}^a - \boldsymbol{x}^0)). \quad (4)
\end{aligned}
$$

This definition describes the minimized ℓ_p-norm (divided by N) under the condition $\boldsymbol{y}(=\boldsymbol{F}\boldsymbol{x}_0) = \boldsymbol{F}\boldsymbol{x}$, which is clearly obtained by taking the limit of $\beta \to \infty$. In the present case, matrix $\boldsymbol{F}$ and input $\boldsymbol{x}_0$ are random, and we need to take the average w.r.t. them. This requires calculating the average of the logarithmic quantity on the rhs, which is an obstacle to the analysis. To overcome this obstacle, we resort to the replica method, which has not been shown to be rigorous but heuristically gives the exact result. With the replica method, C_p after averaging is

$$
\mathbb{E}[C_p]_{\boldsymbol{F},\boldsymbol{x}_0} := -\lim_{\beta \to \infty} \lim_{N \to \infty} \frac{1}{\beta N} \lim_{n \to 0} \frac{\partial}{\partial n} \ln \mathbb{E}[Z^n(\boldsymbol{F},\boldsymbol{x}^0)]_{\boldsymbol{F},\boldsymbol{x}^0},
$$

$$(5)$$

where $\mathbb{E}[\ ]_X$ denotes the average w.r.t. variable (vector, matrix) X; the nth power of factor $Z(\boldsymbol{F},\boldsymbol{x}^0)$ is

$$
Z^n(\boldsymbol{F},\boldsymbol{x}^0) := \prod_{a=1}^{n} \int d\boldsymbol{x}^a \exp(-\beta \|\boldsymbol{x}^a\|_p) \delta(\boldsymbol{F}(\boldsymbol{x}^a - \boldsymbol{x}^0))
$$

$$
= \prod_{a=1}^{n} \int d\boldsymbol{x}^a \lim_{\tau \to +0} \frac{1}{(\sqrt{2\pi\tau})^{nP}}
$$

$$
\times \exp\left[-\sum_{a=1}^{p} \beta \|\boldsymbol{x}^a\|_p - \frac{1}{2\tau} \sum_{a=1}^{n} (\boldsymbol{x}^a - \boldsymbol{x}^0)^T \boldsymbol{F}^T \boldsymbol{F}(\boldsymbol{x}^a - \boldsymbol{x}^0)\right].
$$

$$(6)$$

This means that we can estimate the logarithmic quantity from the positive integer moment with identity $\mathbb{E}[\ln X] = \lim_{n \to 0} \partial \ln \mathbb{E}[X^n]/\partial n$. Superscript a on $\boldsymbol{x}$ denotes the "replica" number introduced for estimating $\mathbb{E}[Z^n(\boldsymbol{F},\boldsymbol{x}^0)]_{\boldsymbol{F},\boldsymbol{x}^0}$. After the average is taken over Gaussian random matrix $\boldsymbol{F}$ and limit $\tau \to +0$,

$$
\mathbb{E}[Z^n(\boldsymbol{F},\boldsymbol{x}^0)]_{\boldsymbol{F}} = \int d\boldsymbol{x}^0 \int d\boldsymbol{Q} \prod_{a=1}^{n} \int d\boldsymbol{x}^a
$$

$$
\times \exp\left(-\frac{\alpha N}{2} \operatorname{Tr} \ln \boldsymbol{S} - \sum_{a=1}^{n} \beta \|\boldsymbol{x}^a\|_p\right) \Pi^{(n)}(\boldsymbol{Q},\boldsymbol{x}^a). \quad (7)
$$

We omit the trivial overall factor as it is irrelevant to the analysis. Square matrix $(\boldsymbol{S})_{ab} := Q_{ab} - 2Q_{0a} + \rho$ is n-dimensional; Q_{ab} is defined in $\Pi^{(n)}(\boldsymbol{Q},\boldsymbol{x}^a)$. Delta function constraint $\Pi^{(n)}(\boldsymbol{Q},\boldsymbol{x}^a)$ is given as

$$
\Pi^{(n)}(\boldsymbol{Q},\boldsymbol{x}^a) :=
$$

$$
\prod_{a=1}^{n} \int_{-i\infty}^{+i\infty} d\widetilde{Q}_{aa} \exp\left\{N\widetilde{Q}_{aa}(\boldsymbol{x}^{aT}\boldsymbol{x}^a - NQ_{aa})\right\}
$$

$$
\times \prod_{a<b} \int_{-i\infty}^{+i\infty} d\widetilde{Q}_{ab} \exp\left\{N\widetilde{Q}_{ab}(\boldsymbol{x}^{aT}\boldsymbol{x}^b - NQ_{ab})\right\}
$$

$$
\times \prod_{a=1}^{n} \int_{-i\infty}^{+i\infty} d\widetilde{Q}_{0a} \exp\left\{N\widetilde{Q}_{0a}(\boldsymbol{x}^{aT}\boldsymbol{x}^0 - NQ_{0a})\right\}. \quad (8)
$$

In this definition, dual matrix $\widetilde{\boldsymbol{Q}}$ is introduced as a collection of integration variables for Fourier representation of the delta function.

Computing the $n \to 0$ limit requires analytic continuation from $n \in \mathbb{N}$ to $n \in \mathbb{R}$. To achieve this, we follow the standard procedure in the replica method and assume replica symmetry regarding matrices $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{Q}}$. Let $q = Q_{ab}, \widetilde{q} = \widetilde{Q}_{ab}$ (both for $a \neq b$), $Q = Q_{aa}, \widetilde{Q} = \widetilde{Q}_{aa}$, $m = Q_{a0}$ and $\widetilde{m} = \widetilde{Q}_{a0}$, yielding $S_{aa} = Q - 2m + u$ and $S_{ab} = q - 2m + u$ ($a \neq b$). By diagonalization of matrix $\boldsymbol{S}$, $\operatorname{Tr} \ln \boldsymbol{S}$ is evaluated as

$$
\operatorname{Tr} \ln \boldsymbol{S} = (n-1)\ln(Q-q) + \ln\{Q-q+n(q-2m+\rho)\}.
$$

$$(9)$$

Under the assumption of replica symmetry, the $\boldsymbol{x}^a$-dependent part, namely the ℓ_p-norm and $\Pi^{(n)}(\boldsymbol{Q},\boldsymbol{x}^a)$ are deformed to

$$
\prod_{a=1}^{n} \int d\boldsymbol{x}^a \exp\left(-\sum_{a=1}^{n} \beta \|\boldsymbol{x}^a\|_p\right) \Pi^{(n)}(\boldsymbol{Q},\boldsymbol{x}^a)
$$

$$
= \exp\left(-NnQ\widetilde{Q} - N\frac{n(n-1)}{2}q\widetilde{q} - Nnm\widetilde{m}\right)
$$

$$
\times \int D\widetilde{z} \left(\int dx \exp\left(N\left\{\left(\widetilde{Q} - \frac{\widetilde{q}}{2}\right)x^2 \right.\right.\right.
$$

$$
\left.\left.\left. + x^T\left(\widetilde{m}x^0 + \sqrt{\widetilde{q}}\widetilde{z}\right) - \beta|x|^p\right\}\right)\right)^n, \quad (10)
$$

where $D\widetilde{z} := (\sqrt{2\pi})^{-1} \int_{-\infty}^{\infty} d\widetilde{z} e^{-\widetilde{z}^2/2}$. In the last line, the interaction between replicas (namely $\boldsymbol{x}^{aT}\boldsymbol{x}^b$) is removed by incorporating auxiliary Gaussian variable $\widetilde{z}$ (often called Hubbard-Stratonovich transformation in physics) and decomposing all replicas.

From (9, 10) we find that all n-dependent factors are taken as defined for $n \in \mathbb{R}$ (putting mathematical rigorousness aside), which allows us to calculate $n \to 0$ limit. For convenience of further analysis, we redefine the auxiliary variables, $\widehat{m} := \beta^{-1}\widetilde{m}, \widehat{\chi} := \beta^{-2}\widetilde{q}, \chi := \beta(Q - q)$, and $\widehat{Q} := \beta^{-1}(-2\widetilde{Q} + \widetilde{q})$, and also introduce the function

$$\phi_p(h, \widehat{Q}) := \frac{1}{N}\min_x\left\{\frac{\widehat{Q}}{2}x^2 - hx + |x|^p\right\}. \quad (11)$$

These variables and function are used to simplify the factor in $\int D\widetilde{z}$ in (10) to $\exp\{-\beta N n\phi_p(\widehat{m}x^0 + \sqrt{\widehat{\chi}}\widetilde{z}, \widehat{Q})\}$ for $\beta \to \infty$.

After combining these results and computing the average w.r.t. x_0, we compute the limit $N \to \infty$. As a result, a six-dimensional integral w.r.t. $\widehat{Q}, \widehat{m}, \widehat{\chi}, q, m, \chi$ is replaced with one w.r.t. their extremal values by asymptotic analysis. This integral is denoted by the symbol Extr in the following. (Although the commutativity of the limits $n \to 0$ and $N \to \infty$ has not been shown, this has not been a concern in standard replica analysis.) Finally, after computing the limit $n \to 0$, we arrive at the final expression:

$$\mathbb{E}[C_p]_{\boldsymbol{F},\boldsymbol{x}^0} = \underset{\substack{\widehat{Q},\widehat{m},\widehat{\chi}\\q,m,\chi}}{\mathrm{Extr}}\left\{\frac{\alpha(q - 2m + u)}{2\chi} + \left(\frac{\chi\widehat{\chi}}{2} - \frac{q\widehat{Q}}{2} + m\widehat{m}\right)\right.$$
$$\left. + \int dx^0 P(x^0)\int D\widetilde{z}\,\phi_p(\widehat{m}x^0 + \sqrt{\widehat{\chi}}\widetilde{z}, \widehat{Q})\right\}. (12)$$

For evaluation of the threshold, we can extract some information from (12). As shown in (12), $\mathbb{E}[C_p]_{\boldsymbol{F},\boldsymbol{x}^0}$ is nothing but the minimized ℓ_p-norm after averaging, which can be calculated using the extremal values of the six variables. Returning to the definition of q, m (delta function in (8)) and remembering that $\boldsymbol{x}$ is the result of recovery and $\boldsymbol{x}^0$ is the original input, we can see that $q = m = \rho$ must be satisfied at the extremal of q, m when the recovery is successful, whereas $q \neq m$ is expected in the unsuccessful case. For threshold evaluation, we need to observe the bifurcation from $q = m$ to $q \neq m$. This statement assumes continuous bifurcation, or second-order phase transition in the context of statistical mechanics, which is true for the present problem.

For observing bifurcation, it is more convenient to use the variable $\chi(= \beta(Q - q))$. In this problem, $Q = q = m$ holds for successful recovery, while $Q \neq q \neq m$ for failure, which means $\chi = 0$ and $\chi \neq 0$ for success and failure, respectively. Henceforth, we use the bifurcation from $\chi = 0$ to $\chi \neq 0$ at the extremal in (12), which yields the recovery threshold in conjunction with the conditions of the other five variables. For $p = 1$ (ℓ_1-norm), the threshold can be expressed as simply two equations:

$$\alpha = 2(1 - \rho)H\left(1/\sqrt{\widehat{\chi}}\right) + \rho,$$
$$\widehat{\chi} = \alpha^{-1}\left\{2(1 - \rho)\left((\widehat{\chi} + 1)H\left(1/\sqrt{\widehat{\chi}}\right)\right.\right.$$
$$\left.\left. - (2\pi)^{-1/2}\sqrt{\widehat{\chi}}e^{-1/2\widehat{\chi}} + \rho(\widehat{\chi} + 1)\right)\right\}, \quad (13)$$

where $H(x) := (2\pi)^{-1/2}\int_x^\infty dt e^{-t^2/2}$ is a complementary error function (slightly different definition from the standard),

and elimination of $\widehat{\chi}$ yields the relation between α and ρ, which is the ℓ_1 recovery threshold. This result coincides with the weak threshold in [4], [5], [6], computed from random polytope projection. (The equivalence is noted in [7]. Actually, the extremal condition w.r.t. ν for cross-polytope in Section 6.2 in [6] is shown to be the same as (13) after some algebra.)

III. UNIVERSALITY FROM STATISTICAL MECHANICAL ANALYSIS: MATRIX INTEGRATION

Numerical investigation of the universality of the ℓ_1-norm recovery threshold [10] using several kinds of i.i.d. random entries in $\boldsymbol{F}$, such as Gaussian and Bernoulli, and several random orthogonal bases, such as Fourier and Hadamard, indicated that the threshold under random matrices is universal.

From a statistical mechanical point of view, this universality is understood by the matrix integration formula [7]. Here we use the formula from Lie group theory [11], which is equivalent to one from mathematical physics [12], called the Harish-Chandra-Itzykson-Zuber integral,

$$\frac{\int d\boldsymbol{O}\exp\left\{\frac{1}{2}\mathrm{Tr}\,\boldsymbol{O}\boldsymbol{D}\boldsymbol{O}^T\boldsymbol{L}\right\}}{\int d\boldsymbol{O}} = \exp\left\{N\,\mathrm{Tr}\,G\left(\frac{\boldsymbol{L}}{N}\right)\right\}, \quad (14)$$

for computing $N \to \infty$, where $\boldsymbol{O}$ is an orthogonal, $\boldsymbol{D}$ is a diagonal, and $\boldsymbol{L}$ is an arbitrary matrix. (This formula was originally given for unitary matrix integration.) All matrices are square and N-dimensional; $d\boldsymbol{O}$ is the Haar measure of the N-dimensional orthogonal group. The function G on the rhs is computed from (see e.g. [13], [14], [15])

$$G(x) = \frac{1}{2}\int_0^x dt\left(\Lambda(t) - \frac{1}{t}\right), \quad (15)$$

which is known as R-transformation in free probability theory [16]. The function $\Lambda(t)$ is implicitly given by Cauchy (or Stieltjes) transformation,

$$x = \int d\lambda \frac{\rho_{\boldsymbol{D}}(\lambda)}{\Lambda(x) - \lambda}, \quad (16)$$

where $\rho_{\boldsymbol{D}}(\lambda)$ is the density of the diagonal element values in $\boldsymbol{D}$.

The random matrix ensemble (Wishart ensemble) $\boldsymbol{F}^T\boldsymbol{F}$, where $\boldsymbol{F}$ is a P-by-N i.i.d. Gaussian random matrix with variance N^{-1}, is assumed to be equivalent to the ensemble $\boldsymbol{O}\boldsymbol{D}\boldsymbol{O}^T$ generated by arbitrary orthogonal matrix $\boldsymbol{O}$ under the condition that $\rho_{\boldsymbol{D}}(\lambda)$ follows Marčenko-Pastur law [17] for $\alpha = P/N < 1$:

$$\rho_{\boldsymbol{D}}(\lambda) = (1 - \alpha)\delta(\lambda)$$
$$+ \frac{1}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}\Theta(\lambda_+ - \lambda)\Theta(\lambda - \lambda_-), (17)$$

where $\lambda_\pm = (1 \pm \alpha^{1/2})^2$ and $\Theta(x)$ is a Heaviside function. (For a unitary ensemble, the equivalence is as shown in [16].) This is known as the asymptotic eigenvalue density of the Wishart random matrix ensemble $\boldsymbol{F}^T\boldsymbol{F}$.

We can apply this formula to the average w.r.t. random matrix $\boldsymbol{F}$ in (6). For (17), the function $G(x)$ is computed as $G(x) = -(\alpha/2)\ln(1-x)$, and applying (14) to (6) (integration

performed over Haar measure $d\boldsymbol{O}$) results in an average the same as that obtained using (7), as noted in [7].

This strongly suggests that rotational invariance of the random matrix ensemble $\boldsymbol{F}^T\boldsymbol{F}$ is combined with the universal threshold because the result of analysis using matrix integration w.r.t. the orthogonal group Haar measure is the same as that using integration of a Gaussian random matrix with i.i.d. entries performed using (6). This implies that universality breakdown requires a random matrix ensemble that breaks such symmetry. In the following, we present a symmetry breaking model and see how the threshold deviates from the universal one.

IV. DEFORMED PROBLEM

As mentioned above, statistical mechanics can be used to analyze problems more general than the basic ℓ_1-norm problem. For example, we previously used it to investigate the deformed problem of an i.i.d. random matrix [8], where $\boldsymbol{F}$ is given by

$$\boldsymbol{F} = \sqrt{\boldsymbol{R}_{\mathrm{r}}}\boldsymbol{\Xi}\sqrt{\boldsymbol{R}_{\mathrm{t}}}. \tag{18}$$

Matrices $\boldsymbol{R}_{\mathrm{r}}$ and $\boldsymbol{R}_{\mathrm{t}}$ are respectively P- and N-dimensional deterministic square symmetric matrices. The square root of square matrix $\boldsymbol{A}$ is defined by $\boldsymbol{A} = \sqrt{\boldsymbol{A}}^T\sqrt{\boldsymbol{A}}$ (e.g. the Cholesky decomposition can be used for positive-definite $\boldsymbol{A}$.). The $\boldsymbol{\Xi}$ is a P-by-N rectangular matrix with entries that are i.i.d. Gaussian random variables with zero mean and variance N^{-1}. As stated elsewhere [8], $\boldsymbol{R}_{\mathrm{r}}$ and $\boldsymbol{R}_{\mathrm{t}}$ respectively describe the correlation among observation vectors and the correlation among the representation bases of the sparse input signals. (Such a framework is described elsewhere [18].) When $\boldsymbol{R}_{\mathrm{r}}$ and $\boldsymbol{R}_{\mathrm{t}}$ are identities, $\boldsymbol{F}$ becomes an i.i.d. random matrix, and the problem returns to the original one.

We previously applied the replica method to this problem and computed $\mathbb{E}[C_p]_{\boldsymbol{F},\boldsymbol{x}^0}$ [8]. The result was

$$\mathbb{E}[C_p]_{\boldsymbol{F},\boldsymbol{x}^0} = \underset{\substack{\widehat{Q},\widehat{m},\widehat{\chi} \\ q,m,\chi}}{\mathrm{Extr}}\left(\frac{\alpha(q-2m+u)}{2\chi} + \left(\frac{\chi\widehat{\chi}}{2} - \frac{q\widehat{Q}}{2} + m\widehat{m}\right)\right.$$
$$\left. + \left\{\prod_i \int dx_i^0 P(x_i^0) \int D\tilde{\boldsymbol{z}}\,\widetilde{\phi}_p(\widehat{m}\sqrt{\boldsymbol{R}_{\mathrm{t}}}\boldsymbol{x}^0 + \sqrt{\widehat{\chi}}\tilde{\boldsymbol{z}},\widehat{Q})\right\}\right), \tag{19}$$

where $\widetilde{\phi}_p(\boldsymbol{h},\widehat{Q})$ is given as the solution to an N-variable minimization problem as

$$\widetilde{\phi}_p(\boldsymbol{h},\widehat{Q}):=\frac{1}{N}\min_{\boldsymbol{x}}\left\{\frac{\widehat{Q}}{2}\boldsymbol{x}^T\boldsymbol{R}_{\mathrm{t}}\boldsymbol{x} - \boldsymbol{h}^T\sqrt{\boldsymbol{R}_{\mathrm{t}}}\boldsymbol{x} + \|\boldsymbol{x}\|_p\right\}. \tag{20}$$

The result clearly does not depend on $\boldsymbol{R}_{\mathrm{r}}$ (assuming $\boldsymbol{R}_{\mathrm{r}}$ is full-rank), and does only on $\boldsymbol{R}_{\mathrm{t}}$. This can be understood from matrix integration. Suppose that $\boldsymbol{R}_{\mathrm{t}}$ is an identity matrix; the ensemble $\boldsymbol{F}^T\boldsymbol{F} = \boldsymbol{\Xi}^T\boldsymbol{R}_{\mathrm{r}}\boldsymbol{\Xi}$ is then equivalent to $\boldsymbol{O}\boldsymbol{D}\boldsymbol{O}^T$ in the previous section because matrix $\boldsymbol{R}_{\mathrm{r}}$ can be eliminated by the redefinition of $\boldsymbol{\Xi}$ (after normalization, which changes the Marčenko-Pastur law of Wishart ensemble $\boldsymbol{\Xi}^T\boldsymbol{\Xi}$ and $G(x)$ in (14), however is irrelevant to the universality [8]). On the other

hand, $\boldsymbol{R}_{\mathrm{t}}$ cannot be eliminated in the same manner and affects the Haar measure $d\boldsymbol{O}$. This implies that such a random matrix ensemble will differ from the one connected from a diagonal matrix like Marčenko-Pastur by orthogonal transformation.

We can calculate the threshold for this deformed problem in a manner similar to that used in the previous section by investigating the bifurcation from $\chi = 0$ to $\chi \neq 0$. Note that, in this problem, we must solve an N-variable minimization problem, as in (20), while in the original problem this is simply a minimization with only one variable. Such minimization generally requires a numerical method (e.g., Monte Carlo) as used previously [8].

We also studied another deformed problem, where input signal $\boldsymbol{x}^0$ is sparse and directly correlated [9]; since this is beyond the scope of this article, we omit the details here.

V. EXAMPLE OF NON-UNIVERSAL THRESHOLD: BLOCKWISE MODEL

For observing in detail the deviation from universality when using the deformed model and statistical mechanics both analytically and quantitatively, we propose using a blockwise model. The sensing matrix in this model is the one given in (18), and $\boldsymbol{R}_{\mathrm{t}}$ is composed of 2-by-2 diagonal blocks. (Input size N is assumed to be even).

$$\boldsymbol{R}_{\mathrm{t}} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \otimes \boldsymbol{I}_{N/2},$$
$$\text{accordingly } \sqrt{\boldsymbol{R}_{\mathrm{t}}} = \begin{pmatrix} l_+ & l_- \\ l_- & l_+ \end{pmatrix} \otimes \boldsymbol{I}_{N/2}, \tag{21}$$

where $l_\pm := (\sqrt{1+r} \pm \sqrt{1-r})/2$, and $\boldsymbol{I}_{N/2}$ denotes an identity matrix of size $N/2$. $\sqrt{\boldsymbol{R}_{\mathrm{t}}}$ is chosen to be symmetric, and $\boldsymbol{R}_{\mathrm{r}} = \boldsymbol{I}_P$. This blockwise matrix represents the case in which all input signals (or equivalently corresponding representation bases) are correlated with their partner.

By substituting these signals or bases into (19) and (20) and observing the bifurcation from $\chi = 0$ to $\chi \neq 0$, we can obtain the recovery threshold for the blockwise model. This analysis scheme is almost the same as that previously proposed [8], so the details are omitted.

The final equations for the threshold are

$$\alpha = \frac{1}{2\sqrt{\widehat{\chi}}}\int Dz_1 \int Dz_2 V(z_1,z_2,\widehat{\chi}),$$
$$\widehat{\chi} = \frac{1}{2\alpha}\int Dz_1 \int Dz_2 W(z_1,z_2,\widehat{\chi}), \tag{22}$$

which corresponds to (13) for the original basic ℓ_1-norm problem. The functions $V(z_1,z_2,\widehat{\chi})$ and $W(z_1,z_2,\widehat{\chi})$ are defined as

$$V(z_1,z_2,\widehat{\chi})$$
$$:= \sum_{\xi_1,\xi_2=0,1}\{(1-\rho)\delta_{\xi_1,0} + \rho\delta_{\xi_1,1}\}\{(1-\rho)\delta_{\xi_2,0} + \rho\delta_{\xi_2,1}\}$$
$$\times \sum_{\sigma_1,\sigma_2=\pm1}\sum_{i,j=1,2}\frac{1}{4(1-r^2)}\left(\sqrt{\boldsymbol{R}_{\mathrm{t}}^B}\right)_{ij}$$
$$\times z_i x_{\xi_1,\xi_2}^{(j)}\left(\sigma_1,\sigma_2,z_1,z_2,\sqrt{\widehat{\chi}}\right), \tag{23}$$

$$W(z_1, z_2, \widehat{\chi})$$
$$:= \sum_{\xi_1,\xi_2=0,1} \{(1-\rho)\delta_{\xi_1,0} + \rho\delta_{\xi_1,1}\}\{(1-\rho)\delta_{\xi_2,0} + \rho\delta_{\xi_2,1}\}$$
$$\times \sum_{\sigma_1,\sigma_2=\pm 1} \sum_{i,j=1,2} \frac{1}{4(1-r^2)} \left(\boldsymbol{R}_{\mathrm{t}}^B\right)_{ij}$$
$$\times x_{\xi_1,\xi_2}^{(i)}\left(\sigma_1, \sigma_2, z_1, z_2, \sqrt{\widehat{\chi}}\right) x_{\xi_1,\xi_2}^{(j)}\left(\sigma_1, \sigma_2, z_1, z_2, \sqrt{\widehat{\chi}}\right), \quad (24)$$

where $\boldsymbol{R}_{\mathrm{t}}^B$ and $\sqrt{\boldsymbol{R}_{\mathrm{t}}^B}$ are the 2-by-2 block matrices in (21). Boolean variables ξ_1 and ξ_2 represent the case in which each input signal is respectively zero and nonzero for pairwise input. The functions $x_{\xi_1,\xi_2}^{(1),(2)}$ for each ξ_1 and ξ_2 are given as

$$x_{0,0}^{(1)}(z_1, z_2, \widehat{\chi}) := \sum_{\eta_1,\eta_2=\pm 1} \Omega_{\eta_1}\left((\widehat{l}_+ z_1 + \widehat{l}_- z_2)\sqrt{\widehat{\chi}} + r\eta_2 - \eta_1\right)$$
$$\times \Theta\left(\eta_2\left\{(\widehat{l}_- z_1 + \widehat{l}_+ z_2)\sqrt{\widehat{\chi}} + r\eta_1 - \eta_2\right\}\right)$$
$$+ \sum_{\eta=\pm 1}(1-r^2)\Omega_\eta\left((l_+ z_1 + l_- z_2)\sqrt{\widehat{\chi}} - \eta\right)$$
$$\times \Theta\left((\widehat{l}_- z_1 + \widehat{l}_+ z_2)\sqrt{\widehat{\chi}} + r\eta + 1\right)$$
$$\times \Theta\left(-\left\{(\widehat{l}_- z_1 + \widehat{l}_+ z_2)\sqrt{\widehat{\chi}} + r\eta - 1\right\}\right),$$
$$x_{0,0}^{(2)}(z_1, z_2, \widehat{\chi}) := \mathrm{replace}\{\widehat{l}_+, \widehat{l}_-, l_+, l_-\} \text{ in}$$
$$x_{0,0}^{(1)}(z_1, z_2, \widehat{\chi}) \text{ with } \{\widehat{l}_-, \widehat{l}_+, l_-, l_+\}, \quad (25)$$

$$x_{0,1}^{(1)}(\sigma_2, z_1, z_2, \widehat{\chi}) := \sum_{\eta=\pm 1} \Omega_\eta\left((\widehat{l}_+ z_1 + \widehat{l}_- z_2)\sqrt{\widehat{\chi}} + r\sigma_2 - \eta\right),$$
$$x_{0,1}^{(2)}(\sigma_2, z_1, z_2, \widehat{\chi}) := -r x_{0,1}^{(1)}(\sigma, z_1, z_2, \widehat{\chi})$$
$$+ (1-r^2)\left\{(l_- z_1 + l_+ z_2)\sqrt{\widehat{\chi}} - \sigma_2\right\}, \quad (26)$$

$$x_{1,0}^{(2)}(\sigma_1, z_1, z_2, \widehat{\chi}) := \sum_{\eta=\pm 1} \Omega_\eta\left((\widehat{l}_- z_1 + \widehat{l}_+ z_2)\sqrt{\widehat{\chi}} + r\sigma_1 - \eta\right),$$
$$x_{1,0}^{(1)}(\sigma_1, z_1, z_2, \widehat{\chi}) := -r x_{1,0}^{(2)}(\sigma, z_1, z_2, \widehat{\chi})$$
$$+ (1-r^2)\left\{(l_+ z_1 + l_- z_2)\sqrt{\widehat{\chi}} - \sigma_1\right\}, \quad (27)$$

$$x_{1,1}^{(1)}(\sigma_1, \sigma_2, z_1, z_2, \widehat{\chi}) := (\widehat{l}_+ z_1 + \widehat{l}_- z_2)\sqrt{\widehat{\chi}} - \sigma_1 + r\sigma_2,$$
$$x_{1,1}^{(2)}(\sigma_1, \sigma_2, z_1, z_2, \widehat{\chi}) := (\widehat{l}_- z_1 + \widehat{l}_+ z_2)\sqrt{\widehat{\chi}} - \sigma_2 + r\sigma_1,$$
$$(28)$$

where $\Omega_\eta(x) := x\Theta(\eta x)$ for $\eta = \pm 1$ and $\widehat{l}_\pm := l_\pm - r l_\mp$. In some cases, $x_{\xi_1,\xi_2}^{(1),(2)}$ does not depend on σ_1 and/or σ_2. Nevertheless, the summations with respect to σ_1 and σ_2 in (23) and (24) are taken in all cases.

This is the main result of this article. Summarizing, we obtain the analytic expression for the ℓ_1-norm recovery threshold by using the two equations in (22) in a similar form as in the original case. The difference is that we need to evaluate a double integral in the present problem, whereas that in the original case is only single integral (in the definition of a complementary error function). When we generalize the deformed problem to the w-blockwise model, we have an
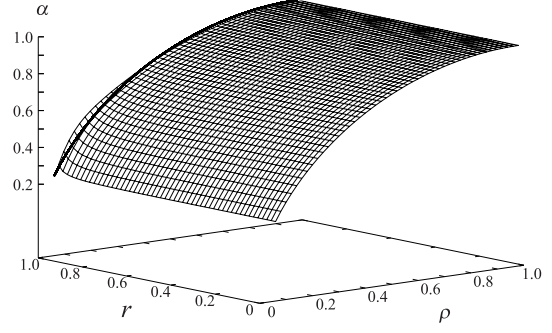

Fig. 1. ℓ_1-norm recovery threshold as function of ρ and r. Areas above and below surface are success and failure regions, respectively. Thick curve at $r = 1$ corresponds to threshold at $r = 0$ (i.e. universal threshold), which is drawn for comparison and for illustrating deviation. In region of large r, a slight deviation from universality is evident.
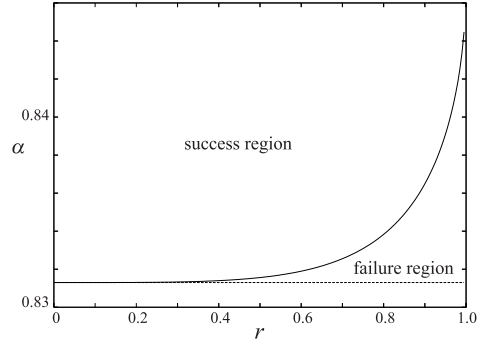

Fig. 2. Dependence of recovery threshold on correlation parameter r for $\rho = 0.5$. Horizontal line corresponds to universal threshold value at $\rho = 0.5$. For large r, deviation from universality is apparent.

expression of the recovery threshold equations with a w-tuple integral. The problem we previously dealt with [8] corresponds to the case in which $w = N$, which requires a Monte Carlo approach to the evaluation of multiple integrals.

Using this analytical expression of the recovery threshold for the deformed problem, we consider the deviation from the universal threshold. The dependence of recovery threshold α on two parameters (ρ and r) is depicted in Fig. 1. In the region where $r \approx 1$, a slight deviation from the universal threshold is evident. The deviation is shown in detail in Fig. 2, where $\rho = 0.5$. In the region of a larger r, a clear deviation from the universal threshold (shown by the horizontal line) is evident.

VI. VALIDATION BY NUMERICAL EXPERIMENT

Given that the replica method lacks rigorousness, we verified its results by conducting a numerical experiment of ℓ_1 recovery. We used the convex optimization package for MATLAB [19], [20] and evaluated the recovery threshold. We first prepared a square random sensing matrix $\boldsymbol{F}$ with size N and deleted the rows in $\boldsymbol{F}$ one-by-one until recovery failure occurred, as determined using $|\boldsymbol{x}_* - \boldsymbol{x}^0| > 10^{-4}$ for recovery result $\boldsymbol{x}_*$. We recorded the number of remaining rows (plus one) $P_c = P + 1$ at the point of failure. We repeated this 10^5 times and then computed the arithmetic average of P_c/N,
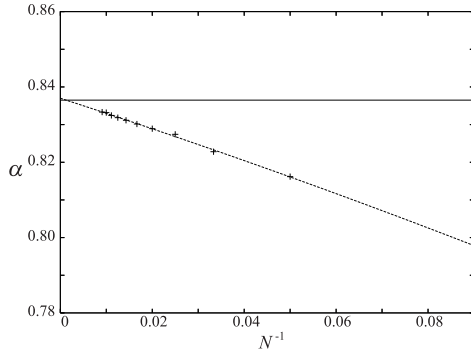
Fig. 3. Result of ℓ_1-norm recovery experiment for $\rho = 0.5$ and $r = 0.9$. Number of input signals N was varied from 20 to 110 in steps of 10. For each N, the average was taken over 10^5 samples. Broken curve indicates scaling by quadratic function regression. Extrapolated value at $N \to \infty$ was 0.8370(3), while result of statistical mechanics evaluated by $N \to \infty$ limit (horizontal line) was 0.83649.... These results are in excellent agreement.

which we regarded as the value of α at the recovery threshold.

The results are plotted in Fig. 3, in which the dependence of the threshold value on the dimension of input signal N is depicted. We also performed scaling analysis using quadratic function regression and estimated the value of the threshold for $N \to \infty$ limit. The value of the threshold from extrapolation was 0.8370(3) for $N \to \infty$. This value is in excellent agreement with the result obtained using statistical mechanics (0.83649...), which validates our analysis for the deformed problem.

VII. Summary and Discussion

We presented a deformed model for ℓ_1-norm reconstruction, that is, a blockwise correlation model that represents the pairwise correlation in signals. From this model we obtained an analytical expression for the ℓ_1-norm recovery threshold by replica heuristics. Using this expression with a double integral, we evaluated the threshold and found a clear deviation from the universal threshold in the region of strong correlation. A numerical experiment validated the results obtained with this model. This model enables minute deviations from the universality of the ℓ_1 recovery threshold to be traced qualitatively by using an analytical expression of the threshold.

We showed that this blockwise model yields a non-universal threshold, as expected from the rotational symmetry breaking argument. This suggests that orthogonality of the representation bases in sensing matrix construction (see [18]) is crucial for universality. The importance of orthogonality to universality was investigated by other researchers in terms of the restricted isometry property [18], [21], and our results support their findings.

The relationship of our analysis to random polytope projection is of great interest. As we noted in Section II, for the original model, both geometrical and statistical mechanical analyses give the same threshold. For the deformed problem, like the one we handled using a blockwise model, the rotational symmetry is broken. It thus appears that a "biased" projection should be taken into account in geometrical analysis, where projection group symmetry, a quotient of rotational group, is generally assumed. The results presented here should be useful for obtaining a deeper understanding of the relationship between geometrical and statistical mechanical analyses.

References

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inform. Theory, vol. 52, no. 2, pp. 489-509, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[3] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," IEEE Trans. Inform. Theory, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[4] D. L. Donoho and J. Tanner, "Neighborliness of randomly projected simplices in high dimensions," Proc. Natl. Acad. Soc., vol. 102, no. 27, pp. 9452-9457, Jul. 2005.

[5] D. L. Donoho, "High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension," Discrete Comput. Geom., vol. 35, no. 4, pp. 617-652, May 2006.

[6] D. L. Donoho and J. Tanner, "Counting faces of randomly projector polytopes when the projection radically lowers dimension," J. Amer. Math. Soc., vol. 22, no. 1, pp. 1-53, Jan. 2009.

[7] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on L_p-norm minimization," J. Stat. Mech., L09003, Sep. 2009.

[8] K. Takeda and Y. Kabashima, "Statistical Mechanical Analysis of Compressed Sensing Utilizing Correlated Compression Matrix," Proc. 2010 IEEE Int. Symp. Inf. Theory, pp. 1538-1542, Jun. 2010.

[9] K. Takeda and Y. Kabashima, "Statistical Mechanical Assessment of a Reconstruction Limit of Compressed Sensing: Toward Theoretical Analysis of Correlated Signals," Europhys. Lett., vol. 95, no. 1, 18006, Jul. 2011.

[10] D. L. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," Phil. Trans. R. Soc. A, vol. 367, no. 1906, pp. 4273-4293, Nov. 2009.

[11] Harish-Chandra, "Differential operators on a semisimple Lie algebra," Amer. J. Math., vol. 79, no. 1, pp. 87-120, Jan. 1957.

[12] C. Itzykson and J. -B. Zuber, "The planar approximation II," J. Math. Phys., vol. 21, no. 3, pp. 411-421, Mar. 1980.

[13] E. Marinari, G. Parisi, and F. Ritort, "Replica field theory for deterministic models: II. A non-random spin glass with glassy behaviour," J. Phys. A, vol. 27, no. 23, pp. 7647-7668, Dec. 1994.

[14] R. Cherrier, D. S. Dean, and A. Lefèvre, "Role of the interaction matrix in mean-field spin glass models," Phys. Rev. E, vol. 67, no. 4, 046112, Apl. 2003.

[15] K. Takeda, S. Uda, and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," Europhys. Lett., vol. 76, no. 6, pp. 1193-1199, Dec. 2006.

[16] A. M. Tulino and S. Verdú, Random Matrix Theory and Wireless Communications, Hanover USA: Now publishers, 2004.

[17] V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," Math. USSR-sbornik, vol. 1, no. 4, pp. 457-483, Apl. 1967.

[18] E. J. Candès and M. B. Wakin, "An Introduction To Compressive Sampling," IEEE Signal Process. Mag., vol. 25, no. 2, pp. 21-30, Mar. 2008.

[19] M. C. Grant and S. P. Boyd, CVX: Matlab software for disciplined convex programming (web page and software) 2009. http://stanford.edu/~boyd/cvx

[20] M. C. Grant and S. P. Boyd, "Graph implementations for nonsmooth convex programs," Recent Advances in Learning and Control, V. Blondel, S. Boyd and H. Kimura eds., pp. 95-110, London:Springer-Verlag, 2008.

[21] R. Baranuik, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," Constr. Approx., vol. 28, no. 3, pp. 253-263, Jan. 2008.